# Direct Search Generalized Simplex Algorithm for Optimizing Non-linear Functions

Hassan Shekarforoush, Marc Berthod, Josiane Zerubia

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

# *Direct Search Generalized Simplex Algorithm for Optimizing Non-linear Functions*

Hassan Shekarforoush, Marc Berthod, Josiane Zerubia

**N° 2535**

Avril 1995

PROGRAMME 4


ISSN 0249-6399

INRIA
SOPHIA ANTIPOLIS

# Direct Search Generalized Simplex Algorithm for Optimizing Non-linear Functions

Hassan Shekarforoush*, Marc Berthod**, Josiane Zerubia***

Programme 4 — Robotique, image et vision
Projet PASTIS

Rapport de recherche n°2535 — Avril 1995 — 23 pages

**Abstract:** Multivariable optimisation techniques have long been used in all fields for improving the design and performance of systems. Yet the number of well known algorithms that can effectively be used under realistic conditions is usually limited due to many practical considerations such as the limit of applicability to certain classes of problems, the time and computational cost of them under conditions of the problem and more importantly, the efficiency of these algorithms under noisy conditions, which is indeed the case in almost all practical problems. Variants of simplex algorithm have been named since 60's as efficient algorithms in noisy situations. However, no theoretical results have been stablished as regards their convergence and computational efficiency. In this report, we have generalized the simplex method and have addressed theoretical aspects concerning the convergence of the algorithm.

**Key-words:** simplex method, search algorithms, optimization.

*(Résumé : tsvp)*

*hshekar@sophia.inria.fr
**berthod@sophia.inria.fr
***zerubia@sophia.inria.fr

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

# Algorithme Généralisé du Simplex pour l'Optimisation de Fonctions Non-linéaires

**Résumé :** Les méthodes d'optimisation multivariable ont été appliquées depuis longtemps dans différents domaines, afin d'améliorer la conception ainsi que les performances des systèmes. Néanmoins, parmi les algorithmes connus, ceux qui peuvent être utilisés dans des conditions réalistes sont limités à cause de la spécificité de ces algorithmes, de leur temps de calcul et surtout de leurs performances en présence du bruit. Depuis les années 60, on s'accorde à affirmer que les variantes de la methode du simplex sont robustes au bruit. Pourtant, on a pu établir des résultats théoriques sur la convergence et l'efficacité de ces algorithmes. Dans ce rapport, nous avons généralisé la méthode du simplex et étudié notamment la convergence de la méthode.

**Mots-clé :** méthode du simplex, algorithmes de recherche, optimisation.

# Table of Contents

# 1 Zero Order Methods

All optimisation methods are based on reconstructing an approximation of the function to be minimised. Optimisation methods can be broadly divided into those which require the gradient of the function and those which do not. Methods depending on the gradient are refered to as the first order or second order methods, depending on whether we need the first or the second gradient of the function (such as the gradient method and the Newton's method). In many practical problems, one can encounter systems that can not be modeled by means of simple mathematical formulae or if so, the model is not easy to handle for computing the first or second derivatives. In such cases the only information available is usually the function values or sample observations (in the case of evolutionary systems) at some intervals. Methods which only employ these information are usually refered to as zero order methods and the simplex algorithm indeed belongs to this class of methods.

A major drawback of first or second order methods is the fact that they are usually very effective, only under conditions close enough to ideal (see for example [8]). This is mainly due to the fact that derivative operators usually amplify noise contents and hence increase the sensitivity of these algorithms. Zero order methods do not suffer from this, at the expense of increase of computational cost. However, unfortunately, most zero order methods, have not been studied thoroughly enough from theoretical points of view.

## 1.1 The Nelder-Mead Version Of The Simplex Algorithm

The Nelder-Mead simplex algorithm [6], which is based on the work of Spendley et al. [12] and Box [1] is an optimization method which attempts to construct a non-local linear approximation of an $n$-dimensional topography from a set of points at sufficient intervals. Although the method has not been investigated thoroughly in theoretical terms, it has been cited as a computationally efficient algorithm for the minimisation of noisy functions and has been used in different areas [7].

The original simplex algorithm proposed by Spendley et al. is based on initializing the algorithm with a regular simplex and moving it iteratively by projecting a vertex through the hyperplane of the remaining vertices in the simplex. The projected point is usually chosen as the vertex where the function to be minimised attains its highest value. The procedure is repeated until a stop criterion is met. Nelder and Mead generalized this algorithm, allowing for changes of shape and size of the simplex. There have been, since then, various studies on the efficiency of the algorithm and its variants [7][10].

The following is a description of the algorithm as proposed by Nelder and Mead in 1965. This algorithm is based on the idea of moving an $n$-dimensional simplex in $\mathbb{R}^n$, using one of four operations described below. Let $x_i^k \in \mathbb{R}^n$, $i = 0 \ldots n$ denote the $n+1$ vertices of a simplex in general positions at iteration k. Let also $f$ denote the function to be minimised in $\mathbb{R}^n$. Then the four operations for moving the simplex are called: reflection, expansion, contraction and the similarity transformation. The algorithm proceeds as follows:

At each iteration, we identify $x_j^k$ for which $f(x_j^k) = max\{f(x_i^k)\}$ and calculate the centroid of the remaining vertices $\bar{x}$. Next we compute the reflected vertex $x_r$:

$$x_r = (1+\alpha)\bar{x} - \alpha x_j^k \tag{1}$$

where $\alpha > 0$ is refered to as the reflection coefficient and is set to 1 in [6]. Note that (1) is equivalent to reflecting $x_j^k$ through the centroid of the opposite side of the simplex and for $\alpha = 1$ the reflection would be symetric with respect to the centroid. Now, if $min\{f(x_i^k)\} < f(x_r) < f(x_j^k)$, $x_j^k$ will be replaced by $x_r$ and we start again with the new simplex. However, if $f(x_r) < min\{f(x_i^k)\}$ we try to see if $x_j^k$ could be moved any further in the same direction. This operation would expand the simplex along the same direction as the reflection and is calculated as below:

$$x_e = \gamma x_r + (1-\gamma)\bar{x} \tag{2}$$

where $\gamma > 1$ is called the expansion coefficient and is set to 2 in [6]. If $f(x_e) < min\{f(x_i^k)\}$ then we replace $x_j^k$ by $x_e$, otherwise the expansion has failed and we replace this vertex by $x_r$.

If on reflecting $x_j^k$ we obtain a new vertex for which $f(x_r) > f(x_i), \forall i \neq j$, then we contract the simplex along the line joining $\bar{x}$ and $x_j^k$:

$$x_c = \beta x_j^k + (1 - \beta)\bar{x} \tag{3}$$

where $0 < \beta < 1$ is the contraction coefficient which is set to 0.5 in [6]. $x_c$ is then accepted unless $f(x_c) > f(x_r)$, in which case the simplex is reduced to one half by a similarity transformation centered at $min\{f(x_i^k)\}$:

$$x_i^{k+1} = \frac{x_i^k + x_l^k}{2} \tag{4}$$

where $f(x_l^k) = min\{f(x_i^k)\}$. The whole procedure is then repeated until a stop criterion is met. Nelder and Mead propose a preset value of $f$ as a stop criterion. Therefore, the procedure is stopped when we hit this value. Figure 1 shows a schematic diagram of a function in $R^2$ and a few iterations of the algorithm up to convergence.

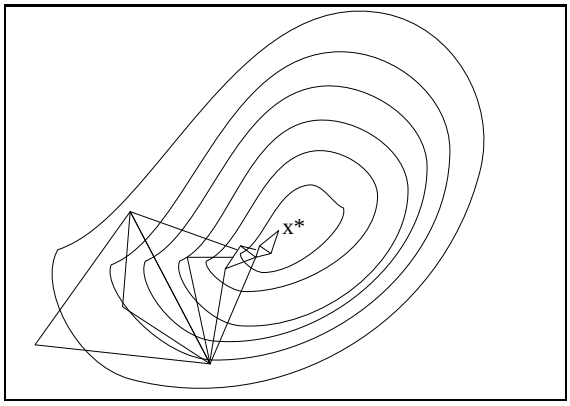


Figure 1: Function level sets and application of the simplex method in $\mathbb{R}^2$

## 1.2 An Overview Of Other Variants Of Simplex Method

There has been many modifications in the simplex algorithm since the publication of Nelder and Mead's powerful adaption of the original algorithm of Spendley et al. [12] (see Parkinson and Hutchinson [7]). There have also been variants of the method proposed by Spendley et al., amongst which we can notably mention the very interesting work of A. S. Rykov [10]. Rykov gives a convergence proof for the case where the simplex is of regular form and under certain restrictive assumtions that the function to be minimised is convex, continuously differentiable and its derivative satisfy the Lipschitz condition:

$$\| \nabla f(x) - \nabla f(y) \| \leq \ell \| x - y \| \tag{5}$$

while the Lebesgue set of the function at any point (see below for definition) is bounded.

A major difference between this version and other published versions is that the author has proposed a method with variable number of reflected vertices. The main argument behind this approach is that in single reflection the angle between the direction along which the centroid of the simplex is moved and the direction of the antigradient of the function may be too large and hence this can reduce the efficiency of the algorithm in terms of the rate of convergence. Therefore, he proposes reflecting a certain optimal number $m$ of worst vertices, which yields a parallel displacement of the reflected vertices along the line joining the centroid of the $m$ vertices and the centroid of the remaining vertices. Therefore, the optimal number of reflected vertices has to be calculated in such a way that the displacement of the centroid of the simplex is as close as possible to the direction of the antigradient at the centeroid. He then accepts a step if the radius of the hypersphere inscribing the simplex does not exceed a threshold value, otherwise the simplex is shrunken by a similarity transformation.

Rykov's algorithm has the drawback that it violates the objective of independency of the algorithm from the gradient and indeed can be considered as a variant of first order methods and hence does not enjoy the same advantages of zero order methods as described above.

# 2 Preliminary Studies

Lack of any theoretical investigation on the simplex method can produce many questions to which one can not put forward a plausible answer. A major problem arising is that we can not determine the class of problems which can be handled by this method. Moreover, one can not determine the qualitative behaviour of the method such as the dependency of the convergence of the algorithm on initial conditions or the nature of the convergence (convergence in function value, in probability, etc.). More importantly, most optimisation algorithms are based on monotonically decreasing a scalar value in an iterative process, which depends on some parameters. The nature and role of these parameters and their range of values usually provide important information on the convergence of the algorithm. A drawback of variants of simplex algorithm is, therefore, in the lack of justification in the choice of values for parameters of the algorithm ($\alpha$, $\beta$ and $\gamma$ in the Nelder-Mead version for instance).

## Theoretical Background

Below, we will develope some theoretical points which will lead to the Generalized Simplex Algorithm with a proof of convergence for convex functions.

**Definition 1** [9] *A scalar function $f(x)$, $x \in \mathbb{R}^n$ is said to be convex if it satisfies the Jensen's inequality. That is $\forall x, y \in \mathbb{R}^n$:*

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) \qquad 0 \leq \lambda \leq 1 \tag{6}$$

*It is said to be strictly convex if:*

$$f((1-\lambda)x + \lambda y) < (1-\lambda)f(x) + \lambda f(y) \qquad 0 < \lambda < 1 \tag{7}$$

*And is said to be strongly convex if:*

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) - \ell\lambda(1-\lambda)\frac{\| x-y \|^2}{2} \qquad 0 \leq \lambda \leq 1 \tag{8}$$

*where $\ell > 0$ is the constant of strong convexity.*

**Definition 2** *Let $f$ be a function defined on $\mathbb{R}^n$. Then the set of points given by:*

$$epif = \{(x, \xi) \mid \xi \geq f(x), x \in \mathbb{R}^n, \xi \in \mathbb{R}\} \tag{9}$$

*is called the epigraph of $f$.*

**Definition 3** *Let $f$ be a function on $\mathbb{R}^n$. Then for any $x_0 \in \mathbb{R}^n$ the set:*

$$\mathcal{L}(x_0) = \{x \in \mathbb{R}^n \mid \mathrm{f(x)} \leq \mathrm{f(x_0)}\} \tag{10}$$

*is called the Lebesgue set [2] of the function $f$ at $x_0$. The set of interior points of $\mathcal{L}(x_0)$ is denoted by:*

$$int\ \mathcal{L}(x_0) = \{x \in \mathbb{R}^n \mid \mathrm{f(x)} < \mathrm{f(x_0)}\} \tag{11}$$

**Remark 1** *It is obvious that the Lebesgue set of a function at any point is always non-empty, since it contains at least the point itself. int $\mathcal{L}(x_0)$ can, however, be empty if the function is bounded below and $x_0 = x^*$, where $x^*$ is the global minimum point. Both $\mathcal{L}(x_0)$ and int $\mathcal{L}(x_0)$ are convex and closed if the function is convex (see [2] and [4]).*

Now, we prove the following statement which we will use later:

**Statement 1** *Let $\xi = f(x)$, $x \in \mathbb{R}^n$ be a convex scalar function and $e = [e_1 \ldots e_{n+1}]$ the standard basis in $\mathbb{R}^{n+1}$. Let also $\pi$ denote a hyperplane parallel to $e_{n+1}$, whose intersection with $f$ is non-empty. Then the cross-section of $\pi$ with $f$ is also a convex function.*

**proof:** Let $\Pi$ denote the set of points in $\pi$. We know that $epif$ is a convex set. On the other hand, $\Pi$ is an affine subspace of $\mathbb{R}^n$. Therefore it is convex, too. Since the intersection of any finite number of convex sets is convex, $epif \cap \Pi$ is also a convex set.

Now, let $\xi = g(x)$ denote the cross-section of $f$ with $\pi$. Take any arbitrary point $[x_1, \xi_1] \in epif \cap \Pi$ for which $\xi_1 = f(x_1) = g(x_1)$, ie. a point that lies on the cross-section of the graph of $f$ with $\pi$. Since $\pi$ is parallel to $e_{n+1}$, for any point $[x_1, \xi_1] \in \Pi$, $[x_1, \xi_2]$, $\xi_2 \neq \xi_1$, is also in $\Pi$. On the other hand, by definition $[x_1, \xi_2]$ is also in $epif$ iff $\xi_2 \geq f(x_1)$ and hence iff $\xi_2 \geq g(x_1)$. Therefore, $[x_1, \xi_2] \in epif \cap \Pi$ iff $\xi_2 \geq g(x_1)$. And this implies that $epif \cap \Pi = epig$. In other words, the epigraph of $g(x)$ is a convex set and therefore, $g(x)$ is also a convex function [3].

**Theorem 1** *Let $f(x)$ be a convex function on $\mathbb{R}^n$. Then for any two distinct points $x, y \neq x^* \in \mathbb{R}^n$, there exists a unique interval $I \subset R$, such that:*

$$f((1-\alpha)x + \alpha y) < min\{f(x), f(y)\} \qquad \forall \alpha \in I \tag{12}$$

**proof:** We first prove (12) in $\mathbb{R}$ and then extend it to $\mathbb{R}^n$.

Without any loss of generality, we will assume that $f(x) \leq f(y)$. Let, now, $x_L$ be an arbitrary point in $int\ \mathcal{L}(x)$, then for $\alpha = \frac{x_L - x}{y - x}$, we have:

$$x_L = (1-\alpha)x + \alpha y \tag{13}$$

It is obvious that $\alpha$ exists since $int\ \mathcal{L}(x) \neq \emptyset$ and $x \neq y$.

On the other hand, let:

$$x_L = (1-\beta)x + \beta y \tag{14}$$

Then:

$$(1-\beta)x + \beta y = (1-\alpha)x + \alpha y \quad \Rightarrow \quad \beta(y-x) = \alpha(y-x) \tag{15}$$

and since $x \neq y$ we can conclude that $\beta = \alpha$.

Therefore, $\alpha$ is unique for any $x_L \in int\ \mathcal{L}(x)$, ie. we have the following one to one affine mapping:

$$\begin{aligned} int\ \mathcal{L}(x) &\longmapsto I \\ x_L &= x - \alpha(x-y) \end{aligned} \tag{16}$$

where $I = \{\alpha \mid \alpha = \frac{x_L - x}{y - x} \ , \ \forall x_L \in int\ \mathcal{L}(x)\}$ is the set containing all values of $\alpha$ corresponding to $x_L \in int\ \mathcal{L}(x)$ for two given distinct points $x$ and $y$ in $\mathbb{R}$.

But, any affine mapping of a convex subset of $\mathbb{R}$ is also convex [4]. On the other hand, any convex set in $\mathbb{R}$ is an interval in $\mathbb{R}$ (the latter follows the definition of a convex set). Therefore, the set $I$ is an interval in $\mathbb{R}$ and is unique for two given distinct points. We have now:

$$(1-\alpha)x + \alpha y \in int\ \mathcal{L}(x) \qquad \forall \alpha \in I \tag{17}$$

Or by definition:

$$f((1-\alpha)x + \alpha y) < min\{f(x), f(y)\} \qquad \forall \alpha \in I \tag{18}$$

This concludes the proof in $\mathbb{R}$.

Now, let $f$ be defined on $\mathbb{R}^n$, and $x$ and $y$ be two distinct points in $\mathbb{R}^n$ with $f(x) \leq f(y)$. Let also $e = [e_1 \dots e_{n+1}]$ denote the standard basis in $\mathbb{R}^{n+1}$. We know that we can pass a unique plane $\pi$ through $x$ and $y$ parallel to $e_{n+1}$. According to Statement 1, the cross-section of $\pi$ with $f$, is a convex function on $\mathbb{R}$ (note that the cross-section is non-empty since it contains $x$ and $y$). On the other hand, since $\pi$ is an affine subspace of $\mathbb{R}^n$, we can treat the two dimensional geometry of $\pi$ directly as a subordinate to the geometry of $\mathbb{R}^n$. Therefore if $g$ represents the cross-section function in $\pi$, we have a unique interval $I \subset R$ for which:

$$g((1-\alpha)x + \alpha y) < g(x) \qquad \forall \alpha \in I \tag{19}$$

But since $\forall x \in \pi$, $g(x) = f(x)$:

$$f((1-\alpha)x + \alpha y) < f(x) \qquad \forall \alpha \in I \tag{20}$$

Or:

$$f((1-\alpha)x + \alpha y) < min\{f(x), f(y)\} \qquad \forall \alpha \in I \tag{21}$$

□

**Corollary 1** *Let $f(x)$ be a convex scalar function on $\mathbb{R}^n$, and $x_0 \dots x_k \in \mathbb{R}^n$, $k < n$ a set of linearly independent vectors such that $f(x_0) = min\{f(x_0) \dots f(x_k)\}$ and int $\mathcal{L}(x_0) \neq \emptyset$. Then there exists a unique interval $I_k \subset \mathbb{R}$, so that:*

$$f((1-\alpha_k)\bar{x}_k + \alpha_k x_k) < f(x_0) \qquad \forall \alpha_k \in I_k \tag{22}$$

*where $\bar{x}_k = \sum_{i=0}^{k-1} \omega_i^k x_i$ and $\omega_i^k$ are some weighting factors given by:*

$$\begin{aligned} \omega_0^1 &= 1 \\ \omega_i^k &= \begin{cases} (1-\alpha_{k-1})\omega_i^{k-1} & \text{if} \quad i = 0 \dots k-2 \\ \alpha_{k-1} & \text{if} \quad i = k-1 \end{cases} \end{aligned} \tag{23}$$

**proof:** We will prove (22) by induction.

for $k = 1$, there exists a unique interval $I_1$ so that:

$$f((1-\alpha_1)\omega_0^1 x_0 + \alpha_1 x_1) < f(x_0) \quad , \quad \forall \alpha_1 \in I_1 \tag{24}$$

This is true according to theorem 1.

Let (22) be also true for $k = m-1 < n-1$, ie. for a unique interval $I_{m-1}$:

$$f((1-\alpha_{m-1}) \sum_{i=0}^{m-2} \omega_i^{m-1} x_i + \alpha_{m-1} x_{m-1}) < f(x_0) \quad , \quad \forall \alpha_{m-1} \in I_{m-1} \tag{25}$$

Take: $y = (1-\alpha_{m-1}) \sum_{i=0}^{m-2} \omega_i^{m-1} x_i + \alpha_{m-1} x_{m-1}$

$y$ and $x_m$ are distinct vectors since $x_0 \dots x_{m-1}, x_m$ are assumed to be linearly independent. Therefore, according to theorem 1, there exists a unique interval $I_m$ so that:

$$f((1-\alpha_m) y + \alpha_m x_m) < f(y) < f(x_0) \quad , \quad \forall \alpha_m \in I_m \tag{26}$$

Or:

$$f((1-\alpha_m)((1-\alpha_{m-1}) \sum_{i=0}^{m-2} \omega_i^{m-1} x_i + \alpha_{m-1} x_{m-1}) + \alpha_k x_m) < f(x_0) \quad , \quad \forall \alpha_m \in I_m \tag{27}$$

Simplifying using the recursive values in (23), this is equivalent to:

$$f((1-\alpha_m) \sum_{i=0}^{m-1} \omega_i^m x_i + \alpha_m x_m) < f(x_0) \quad , \quad \forall \alpha_m \in I_m \tag{28}$$

□

We can show the following properties:

**P- 1**

$$\sum_{i=0}^{k-1} \omega_i^k = 1 \quad , \quad \forall k < n \tag{29}$$

**P- 2** *$I_k$ satisfies one of the following:*

$$\begin{cases} I_k \subset ]0,1[ \\ \text{or} \\ I_k \subset ]-\infty,0[ \end{cases} \tag{30}$$

P-1 can be easily shown by induction:

It is obvious that for $k = 1$, (29) is true.
Let for $k = m - 1$:

$$\sum_{i=0}^{m-2} \omega_i^{m-1} = 1 \tag{31}$$

Then:

$$(1 - \alpha_{m-1}) \sum_{i=0}^{m-2} \omega_i^{m-1} + \alpha_{m-1} = 1 \tag{32}$$

Using the recursive values in (23):

$$\sum_{i=0}^{m-1} \omega_i^m = 1 \tag{33}$$

In order to establish P-2, we first need to recall the reverse Jensen's inequality and the notion of the subgradient (note that a convex function, in general, is not necessarily differentiable):

**Lemma 1** (Reverse Jensen's Inequality) *Let $f(x)$ be a convex function on $\mathbb{R}^n$. Then for any two arbitrary points $x$, $y \in \mathbb{R}^n$:*

$$f((1-\alpha)x + \alpha y) \geq (1-\alpha)f(x) + \alpha f(y) \quad , \quad \forall \alpha \notin ]0,1[ \tag{34}$$

For proof see [2].

**Definition 4** *Let $f(x)$ be a convex function on $\mathbb{R}^n$. Then any vector $\partial f(x) \in \mathbb{R}^n$ for which:*

$$f(x+y) \geq f(x) + \langle \partial f(x), y \rangle \quad , \quad x, y \in \mathbb{R}^n \tag{35}$$

*is called a subgradient of $f$ at the point $x$. $\langle \partial f(x), y \rangle$ denotes the scalar product of the two vectors.*

From the continuity of $f$ [8], we know that there exists at least one such a vector at any point. However, this vector is not necessarily unique everywhere.

**Remark 2** *The convexity of a function in $\mathbb{R}^n$ is equivalent to the monotonicity of the subgradient* [8]*:*

$$\langle \partial f(x) - \partial f(y), x - y \rangle \geq 0 \tag{36}$$

*For a convex function on $\mathbb{R}$ this implies that $\partial f(x) \geq \partial f(y)$ for $x \geq y$.*

P-2 can, now, be verified as follows: it is obvious that $1 \notin I_k$. Therefore, from the uniqueness of the interval $I_k$ we can deduce that $\alpha_k$ is either smaller or greater than 1.

On the other hand, according to (34) if $\alpha_k > 1$ then:

$$f((1-\alpha_k)\bar{x}_k + \alpha_k x_k) \geq (1-\alpha_k) f(\bar{x}_k) + \alpha_k f(x_k) \tag{37}$$

Assuming that $\alpha_k = 1 + \epsilon_k$ where $\epsilon_k > 0$, we get:

$$f((1-\alpha_k)\bar{x}_k + \alpha_k x_k) \geq -\epsilon_k f(\bar{x}_k) + (1+\epsilon_k) f(x_k) \tag{38}$$

And since according to corollary 1 $f(x_k) > f(\bar{x}_k)$:

$$f((1-\alpha_k)\bar{x}_k + \alpha_k x_k) > f(x_k) \tag{39}$$

In other words, for $\alpha_k > 1$ corollary 1 is not satisfied. Therefore, $I_k$ is always strictly below 1.

Now, if $I_k$ contains both positive and negative values, then we can find a positive $\alpha_k \in I_k$ so that $-\alpha_k$ is also in $I_k$. But then since both $\alpha_k$ and $-\alpha_k$ are in $I_k$, according to Theorem 1:

$$f((1-\alpha_k)\bar{x}_k + \alpha_k x_k) < f(\bar{x}_k) \tag{40}$$
$$f((1+\alpha_k)\bar{x}_k - \alpha_k x_k) < f(\bar{x}_k) \tag{41}$$

Using the definition of the subgradient:

$$\langle \partial f(\bar{x}_k), \alpha_k(\bar{x}_k - x_k)\rangle \leq f((1-\alpha_k)\bar{x}_k + \alpha_k x_k) - f(\bar{x}_k) \tag{42}$$
$$\langle \partial f(\bar{x}_k), -\alpha_k(\bar{x}_k - x_k)\rangle \leq f((1+\alpha_k)\bar{x}_k - \alpha_k x_k) - f(\bar{x}_k) \tag{43}$$

And hence from (40) and (41):

$$\langle \partial f(\bar{x}_k), \alpha_k(\bar{x}_k - x_k)\rangle < 0 \tag{44}$$
$$\langle \partial f(\bar{x}_k), -\alpha_k(\bar{x}_k - x_k)\rangle < 0 \tag{45}$$

which is impossible. Therefore, such $\alpha_k$ does not exist, ie. we can not have both positive and negative values for $\alpha_k$.
A geometric interpretation of P-2 is as follows:

Let $\bar{x}_k$ and $x_k$ be the position vectors of the two points $\bar{X}_k$ and $X_k$ in $\mathbb{R}^n$. Let also $x_\alpha$ denote the position vector of the point given by the convex combination $(1-\alpha_k)\bar{x}_k + x_k$, which lies along the line $L_k$ passing through $\bar{X}_k$ and $X_k$. Corollary 1 is, merely, reducing an $n$-dimensional problem into a one dimensional one, ie. a point $X_\alpha$ is sought in such a way that will reduce the value of the function $f$ along the line $L_k$. This is done by passing a plane through $\bar{X}_k$ and $X_k$ parallel to $e_{n+1}$ and reducing the function $f$ to a two dimensional function $g$ (see Figure 2 for example).

Recall that $g$ is a convex function on $\mathbb{R}$. Therefore, from the monotonicity of the subgradient of a convex function, starting from the point $X_k$, $g$ can only reduce along one direction on the line $L_k$ and not both and since $g(x) = f(x)$ for any $x$ in the plane $\pi$, the same applies to $f$. Therefore, P-2 simply substantiates that $X_\alpha$ is always on a half line which contains the point $\bar{X}_k$ and is given by $x_\alpha$ for $\alpha < 1$. Figures 3 and 4 show some examples of $g$ and different possible values for $\alpha_k$.

Note that $x_\alpha$ is the sum of the two vectors $\bar{x}_k$ and $\alpha_k(x_k - \bar{x}_k)$. Therefore, it is easy to see that for the two examples given in Figure 3, $\alpha_k$ has to be negative so that

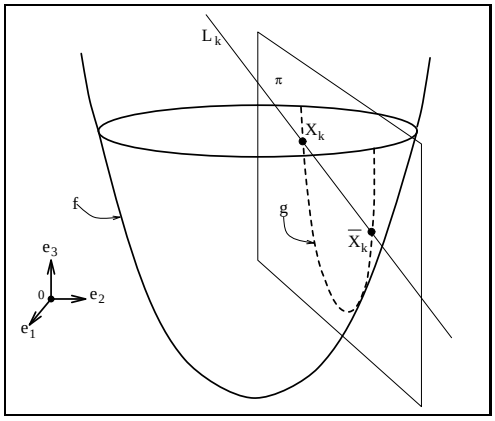


Figure 2: An example of the cross-section of a convex function $f$ defined on $\mathbb{R}^2$ with a plane parallel to $e_3$

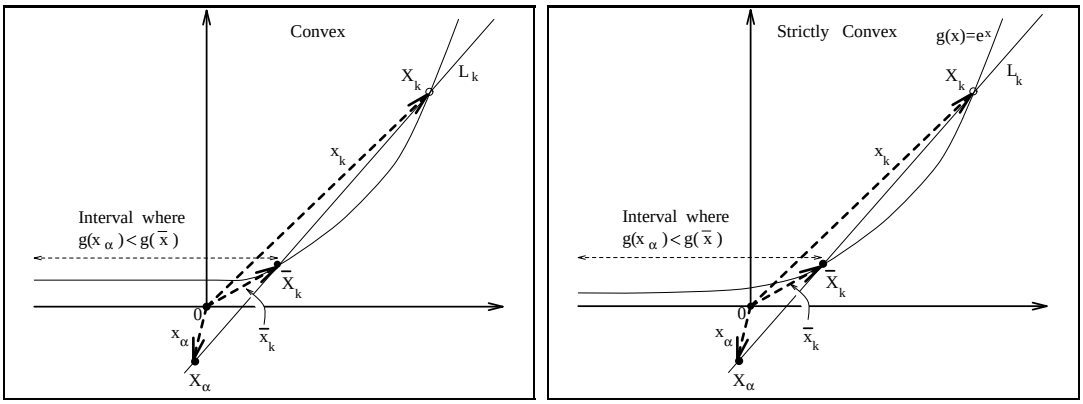


Figure 3: A convex and a strictly convex $g(x)$ with $\alpha_k < 0$

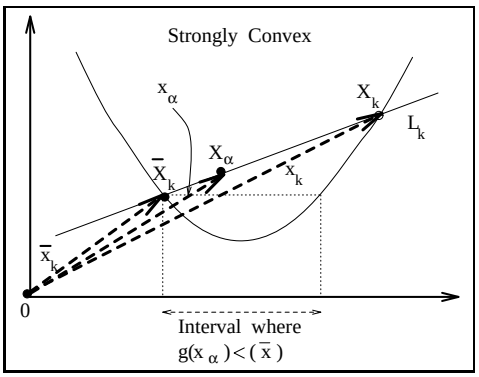


Figure 4: A strongly convex $g(x)$ with $0 < \alpha_k < 1$

$X_\alpha$ can be in the interval where $f(x_\alpha) < f(\bar{x})$, whereas in Figure 4, $\alpha_k$ is certainly in $]0, 1[$.

Let us, now, return back to $\alpha_k$. In general, it is not possible to specify the entire interval $I_k$ to which $\alpha_k$ belongs. However, by taking more restrictive assumptions we can specify a subinterval of $I_k$.

For the case when $0 < \alpha_k < 1$, we can impose the more restrictive assumption of:

$$(1 - \alpha_k) f(\bar{x}_k) + \alpha_k f(x_k) < f(x_0) \tag{46}$$

Then according to Jensen's inequality, (22) is automatically satisfied and we obtain:

$$\alpha_k < \frac{f(x_0) - f(\bar{x}_k)}{f(x_k) - f(\bar{x}_k)} \tag{47}$$

For $\alpha_k < 0$ we have:

$$f((1 - \alpha_k)\bar{x}_k + \alpha_k x_k) \geq (1 - \alpha_k) f(\bar{x}_k) + \alpha_k f(x_k) \tag{48}$$

$$2f(\bar{x}_k) > 2f((1 - \alpha_k)\bar{x}_k + \alpha_k x_k) \tag{49}$$

where (48) follows the reverse Jensen's inequality and (49) follows Theorem 1.

Adding both sides and simplifying:

$$f((1-\alpha_k)\bar{x}_k + \alpha_k x_k) < (1+\alpha_k)f(\bar{x}_k) - \alpha_k f(x_k) \tag{50}$$

Now, again under the more restrictive assumption of $(1+\alpha_k)f(\bar{x}_k) - \alpha_k f(x_k) < f(x_0)$, we can guaranty that $f((1-\alpha_k)\bar{x}_k + \alpha_k x_k) < f(x_0)$. And hence, we will get:

$$\alpha_k > \frac{f(\bar{x}_k) - f(x_0)}{f(x_k) - f(\bar{x}_k)} \tag{51}$$

# 3 The Generalized Simplex Algorithm

The foregoing discussion inspires an optimization algorithm which exploits properties of simplices in $\mathbb{R}^n$. The algorithm is given below in Definition 8:

**Definition 5** *Let $x_0 \dots x_n$ be a set of points in $\mathbb{R}^n$. Take any one of these points as the origin (eg. $x_j$). Then the remaining points are said to be linearly independent if the matrix $(x_0 - x_j \dots x_n - x_j)$ has rank $n$.*

Geometrically, this implies that the hyperplane passing through $x_i$ , $i \neq j$ will yield:

$$\langle a, x_i \rangle + b = 0 \quad , \quad a \in \mathbb{R}^n \, , \; b \in \mathbb{R} \tag{52}$$

which is a set of homogeneous linear equations in $a$ and $b$ whose coefficient matrix $(x_0 \dots x_n)$ , $i \neq j$ has rank $n$. This implies that the smallest linear subspace of $\mathbb{R}^n$ containing the points $x_i$ , $i \neq j$ is $\mathbb{R}^n$ itself.

Alternatively, the vectors $x_0 - x_j \dots x_n - x_j$ are said to be linearly independent if $\sum_{i \neq j} \gamma_i (x_i - x_j) = 0$ implies that $\gamma_i = 0$, $\forall i = 0 \dots n$ , $i \neq j$.

Points satisfying the above condition will be, hereafter, refered to as being in general positions in $\mathbb{R}^n$.

**Definition 6** *Let $x_0 \dots x_n$ be $n+1$ points in general positions in $\mathbb{R}^n$. Then all points $\sigma$ in $\mathbb{R}^n$, which satisfy:*

$$\begin{cases} \sigma = (x_0 - x_j \dots x_n - x_j)\omega \quad , \quad \omega \in \mathbb{R}^n \\ \sum_{i=0}^{n} \omega_i = 1 \quad , \quad i \neq j \end{cases} \tag{53}$$

*form a span of* $x_0 - x_j \dots x_n - x_j$ *which we will refer to as a quasi-euclidean n-simplex in* $\mathbb{R}^n$. $\omega_i$ *are the elements of the vector* $\omega$ *and are the barycentric coordinates relative to* $x_0 - x_j \dots x_n - x_j$. $x_0 \dots x_n$ *are called the vertices of the simplex.*

*The points* $\sigma$ *form a euclidean n-simplex if* $\omega_i \geq 0$, $\forall i$. *The euclidean k-simplex is in fact the smallest convex hull of* $x_0 \dots x_n$.

Unless otherwise mentioned, a quasi-euclidean $n$-simplex in $\mathbb{R}^n$ will be, hereafter, briefly refered to as an $n$-simplex in $\mathbb{R}^n$. Moreover, for a given function on $\mathbb{R}^n$, we shall use $x_l$ and $x_h$ for denoting the vertices where $f(x_l) = min\{f(x_i)\}$ and $f(x_h) = max\{f(x_i)\}$ , $\forall i = 0 \dots n$.

**Definition 7** *Let* $f(x)$ *be a function defined on* $\mathbb{R}^n$. *Let also* $f(x)$ *be bounded below with* $f(x^*) = minf(x)$ , $\forall x \in \mathbb{R}^n$. *Then a sequence* $x^t \in \mathbb{R}^n$, $t \in \mathbb{N}$, *is said to be a minimizing sequence for* $f$ *(or briefly a minimizing sequence) if:*

$$f(x^t) \to f(x^*) \tag{54}$$

**Definition 8** (*Generalised Simplex Algorithm*) *Let* $f(x)$ *be a convex scalar function on* $\mathbb{R}^n$. *We define the Generalised Simplex Algorithm as the iterative process* $\sigma^{t+1} = \mathcal{T}(\sigma^t)$, *where* $\sigma^t$ *is an n-simplex in* $\mathbb{R}^n$ *at iteration t and* $\sigma^{t+1}$ *is an n-simplex which differs from* $\sigma^t$ *in only one vertex. The modified vertex is obtained by replacing* $x_h^t$ *by* $(1 - \alpha_t)\bar{x}_t + \alpha_t x_h^t$.

It is obvious that under the same assumptions as in corollary 1:

$$(1 - \alpha_t)\bar{x}_t + \alpha_t x_h = x_l^{t+1}$$

Therefore, $f(x_l^t)$ is a strictly decreasing sequence and since $f$ is bounded below it is also convergent.

One can also show that $\sigma^t$ is a non-degenerate $n$-simplex ie. at each iteration, the vertices remain in general positions in $\mathbb{R}^n$. Let the vertices be in general positions at iteration $t$. Then at iteration $t+1$ we have the following vertices:

$$x_l^{t+1}, x_i^t \quad , \quad \forall i \neq h \tag{55}$$

Now let:

$$\sum_{i \neq h} \gamma_i (x_i^t - x_l^{t+1}) = 0 \quad , \quad \gamma_i \in \mathbb{R} \tag{56}$$

According to Corollary 1: $x_l^{t+1} < min\{x_i^t\}$ , $\forall i$. Therefore, $x_i^t - x_l^{t+1} \neq 0$ which implies that $\gamma_i = 0$. Therefore, we conclude that $\sigma^{t+1}$ is non-degenerate.

Let, now:

$$f(x_l^t) \rightarrow f(\hat{x}_l) \text{ so that } f(\hat{x}_l) \neq f(x^*) \tag{57}$$

Then since $\hat{x}_l$ belongs to a non-degenerate $n$-simplex and $int\ \mathcal{L}(\hat{x}_l)$ is non-empty, there exists some $\alpha$ so that $f((1-\alpha)\bar{x} + \alpha\hat{x}_h) < \hat{x}_l$. But this contradicts (57). Therefore: $f(\hat{x}_l) = f(x^*)$, ie. $x_l^t$ is a minimising sequence for $f$.

**Definition 9** *A global minimum point is defined to be globally stable if any minimising sequence converges to it* [8]. *That is for any minimising sequence $x^t$:*

$$f(x^t) \rightarrow f(x^*) \quad \Rightarrow \quad x^t \rightarrow x^* \tag{58}$$

**Definition 10** *The set $X^*$ of global minimum points of a function $f$ is said to be weakly stable if any minimising sequence converges to a point in $X^*$* [8]*:*

$$f(x^t) \rightarrow f(x^*) \quad \Rightarrow \quad x^t \rightarrow x^*, \ \ x^* \in X^* \tag{59}$$

**Theorem 2** *The unique minimum point of a strictly (strongly) convex function on $\mathbb{R}^n$ is globally stable and the set of minimum points $X^*$ of a convex function on $\mathbb{R}^n$ is weakly stable if its Lebesgue set is bounded for all points in $\mathbb{R}^n$.*

For proof see [8]

**Remark 3** *The iterative process $\mathcal{T}$ defined in Definition 8 will always converge to one of the global minimum points of $f$ if $f$ is convex with bounded Lebesgue sets and to the unique global point of $f$ if it is stricly (strongly) convex.*

This completes the proof of the convergence of the Generalized Simplex Algorithm for convex functions. As for the rate of convergence, according to Theorem 2:

$$\lim_{t\to\infty} \frac{\| x_0^{t+1} - x^* \|}{\| x_0^t - x^* \|} < 1 \tag{60}$$

This implies that the convergence is linear [5].

## Initialisation

Recall that for the initial simplex we have only required that it should be general, ie. any $n$ vertices of the simplex should be linearly independent, if we take the $n+1$th vertex as the origin. Note that the maximum number of possible linearly independent vectors in $\mathbb{R}^n$ is $n$ [11].

Taking any one of these points as the origin (eg. $x_0$), one can then express the remaining points by:

$$x_i = x_0 + \delta e_i \quad , \quad i = 1 \ldots n \tag{61}$$

where $e_i$ are $n$ unit vectors that are linearly independent and $\delta$ is the radius of the hypersphere inscribing the simplex. It is, therefore, possible to control the size of the simplex by means of this parameter.

To initialize the algorithm we need to initialize the parameters $\omega_0^0 \ldots \omega_{n-1}^0$. This is done as follows: we first generate $n$ linearly independent vectors (using for example sequences of Gram determinants [11]) . We then generate two random vectors. Amongst all these, we identify the vectors which yield the smallest and the highest values for our objective function $f$ (denote these vectors by $\bar{x}$ and $x_h$, respectively). $\omega_0^0 \ldots \omega_{n-1}^0$ are then found by solving the following linear equation:

$$X\omega = \bar{x} \tag{62}$$

where $X$ is an $n \times n$ matrix whose columns are the above generated vectors except $\bar{x}$ and $x_h$. $\omega$ is the vector: $[\omega_0^0 \ldots \omega_{n-1}^0]^T$, $T$ denoting the transposition.

Once (62) is solved, we obtain $n$ linearly independent vectors which are given by the columns of $X$ and their corresponding $\omega_i$ factors. These together with $x_h$ provide the initialization of the algorithm while (47) and (51) are used for iterating.

# 4 Conclusion

Different versions of the simplex algorithm proposed in the literature, have been unified, herein, in a single generalized version. A rigorous proof of convergence has been given for convex functions. The rate of convergence has been established as being linear, which is comparable to the gradient method. Tests are being carried out with vector variables of 400 elements (particularly in the area of image processing).

# References

[1] G. E. P. Box. Evolutionary operation: A method for increasing industrial productivity. In *International Conference On Statistical Quality Control*, 1955.

[2] V. F. Demyanov and L. V. Vasilev. *Nondifferentiable Optimization.* Optimization Software, Inc., 1985.

[3] R. V. Gamkrelidze. *Encyclopaedia Of Mathematical Sciences: Analysis II*, volume 14. Springer-Verlag, 1990 (Russian version 1987).

[4] B. Grünbaum. *Convex Polytopes.* Interscience Publishers, 1967.

[5] D. G. Luenberger. *Linear And Nonlinear Programming.* Addison-Wesley Publishing Company, reprinted 2nd edition, 1989.

[6] J. Nelder and R. A. Mead. A simplex method for function minimization. *The Computer Journal*, 7, 1965.

[7] J. M. Parkinson and D. Hutchinson. An investigation into the efficiency of variants of the simplex metod. In Lootsma F. A., editor, *Numerical Methods For Non-linear Optimization.* Academic Press, 1972.

[8] B. T. Polyak. *Introduction To Optimization.* Optimization Software, Inc., 1987.

[9] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, 1972.

[10] A. S. Rykov. Simplex direct search algorithms. *Automation And Remote Control*, 41, 1980.

[11] L. L. Scharf. *Statistical Signal Processing.* Addison-Wesley Publishing Company, 1990.

[12] W. Spendley, G. R. Hext, and F. R. Himsworth. Sequential applications of simplex designs in optimisation and evolutionary operation. *Technometrics*, 4(4), 1962.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399